Measuring Ideology on the Courts*

Michael A. Bailey
Department of Government and McCourt School of Public Policy
Georgetown University
Michael.Bailey@georgetown.edu

June 2016

Abstract

Measuring ideology on the courts is important, yet challenging. This paper discusses various approaches, ranging from calculating percent conservative scores to measures based on sophisticated measurement models, emphasizing two themes. First, scholars should not assume that all ideological measures work in all research contexts. I identify specific instances in which sophisticated measures produce questionable results. Second, ideology on the court is distinctive from ideology among elected officials, particularly on the Supreme Court not only because justices are influenced by legal factors, but also because the small size of the Court makes idiosyncratic world views of justices (and median justices in particular) decisive. The result is that the court's decisions may lack the ideological coherence more typically observed in Congress. While the task of measuring ideology on the courts may seem technical, it is in fact highly politically relevant and quite unsettled. The final part of this paper explores the competing and evolving measures of ideology for the Roberts Court.

*Draft prepared for 2016 Southern Political Science Association Meetings, January 2016. I'm grateful for helpful comments from Joshua Fischman, John Kastellec and Matt Spitzer.

Measuring ideology of the courts is important. Scholars cannot test some of the most interesting theories of court behavior without some way to summarize the ideological orientation of the justices. Pundits, politicians and observers of the Court have a huge appetite for characterizing the court's ideology, whether to decry justices' extremism or laud their even-handedness.

Measuring judicial ideology is not easy, however. Conceptually, we need to answer several knotty questions before going to the data. Is ideology simply a summary of observed behavior? Is it different from law? Is it one dimensional? Does it vary over time for individual justices? Empirically, scholars have produced a range of approaches, each with its strengths and none without its challenges.

This paper describes the approaches scholars have used to measure judicial ideology. One important theme is the critical need to link the measure used to the needs of an application. Sometimes, simple suffices. Other times, sophistication is essential. Sophistication alone is never sufficient, however, as even some of the most complex approaches may not serve research requirements in many cases. I show specific examples in which sophisticated measurement models produce results that are not credible or consistent.

A second theme is that the distinction between law and ideology can be murky for the Court. Justices are very, but not simply, ideological (Solum 2014). There are twists and turns in their thinking that defy simple categorization. Often we may, for very good analytical and practical reasons, force judicial ideology into pre-defined political ideological space, but we should not forget that ideology is not easily distinguished from law.

This paper proceeds from simple to complex models and finishes with a discussion of the

Roberts Court. Part One discusses the foundations of ideological measurement by presenting approaches that are relatively transparent and often perfectly serviceable. Part Two discusses one-dimensional models based on item response theory (IRT). These models do most of the measurement work in contemporary research, but nonetheless have challenges that sometimes go unappreciated. Part Three explores various approaches to identifying preference differences across time and among institutions. Part Four discusses multidimensional models, models that more accurately characterize court behavior, but can blur the distinction between ideology and law. Part Five uses differing views about the conservatism of the Roberts Court as a lens through which to view the complexity (and, dare I say, excitement?) of measuring the ideology of the Supreme Court. Part Six concludes.

1 Foundations

No serious discussion of the modern courts can avoid ideology. Pritchett (1948) focused on the ideological divisions on the court, highlighting the "judiciary as a participant in the political process, subject to study within the same frame of reference as other governmental institutions" (Pritchett 1970, 983). Segal and Spaeth (1994; 1996; 2001; 2002) have made compelling arguments regarding the primacy of ideology on the Supreme Court.

The easiest way to measure the ideology of Supreme Court justices is simply to tally how often they vote conservatively. This information can be readily extracted from the Supreme Court Database (Spaeth et al 2015). While using percent conservative as a preference measure has largely been superseded by the measurement approaches discussed below, it is not

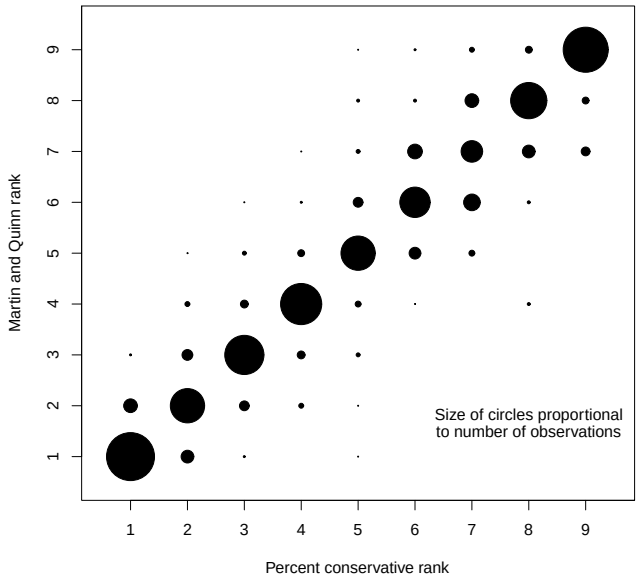

Figure 1: Ordinal Martin and Quinn Scores versus Percent Conservative

without its advantages.[1] First, it is easy. One simply applies basic math to readily available data to produce ideological measures that make sense.

Second, percent conservative scores perform well for some measurement contexts. If one simply needs an ordinal measure of justice ideology for each term, percent conservative scores produce rankings that look very similar to rankings of more complicated measurement models. For example, Figure 1 shows a scatterplot of ranks produced by percent conservative scores and by the Martin and Quinn IRT model discussed below. The size of the circles reflects how many observations are at a given point and the results illustrate how similar the ordinal rankings are. In fact, the ordinal ranking from Martin and Quinn and percent conservative scores correlate at 0.93 for the period 1950 to 2011.

Nonetheless, the disadvantages of percent conservative scores are clear. One is that the approach requires a coding of liberal and conservative votes for each Supreme Court case.

[1] Much of the literature uses percent liberalism. Obviously percent conservatism is a simple function of percent liberalism. It offers the advantage of putting conservatives on the right, something that helps simplify discussions when ideological measures are graphed.

Spaeth (2015) uses rule-based coding, deeming a decision liberal if the decision favors an accused criminal, a civil rights claimant (typically minorities, homosexuals, poor people or some other "underdog"), the government in most takings cases, a free speech or privacy claim and so on. Usually Spaeth coding is unproblematic, but not always. Sometimes Spaeth coding is hinky. On campaign finance cases, for example, Spaeth codes voting to uphold the 2004 Bipartisan Campaign Reform Act (BCRA) as conservative because they limited speech even though this was the politically liberal position.[2]

One trick to coding the ideologically of case is to let politicians and interest groups do the work. Harvey and Woodruff (2011), Bailey (2013) and Hansford (2012) use auxiliary political information to identify which side is liberal and which is conservative. For BCRA, for example, the vote in Congress made it clear that the political valence of upholding BCRA was liberal. Such coding can be laborious and potentially subjective in some instances, although no more subjective than the coding of the Supreme Court database.

Some cases resist simple ideological coding, a fact relevant as we think about ideology and the court. In *Gonzales v. Raich* (2005) the Court overturned a California law allowing medical marijuana by reasoning the federal law superseded state law. Was *Raich* liberal because it favored a federal law over a state law, something liberals typically favor? Or was it a conservative decision because it ruled against medical marijuana? Such cases are relatively rare, but hardly unheard of. How much effort one undertakes to get them right depends on the point of the research. If the point of the research is to characterize the Court's conservativism in general terms, such cases can be relegated to "error" without

[2]Harvey and Woodruff (2011) note that coders may code according to what they think they directionality "should" be based on how the vote turned out, introducing potential distortions.

risking too much bias or imprecision. However, if the purpose is to explain specific cases or to characterize the nuances of individual justice's approach to deciding cases, then one may wish to expend considerable effort in coding and modeling them appropriately as did Shapiro (2009, 2010). The multidimensional approaches we discuss below are also relevant for such cases.

Another concern with percent conservative scores is that they depend on which cases a justice voted on. In one year, the court might have more cases where even conservative justices vote liberally and the next year may feature more cases where even liberal justices vote conservatively. The percent conservative numbers will bounce around depending on such changes in the caseload, even if there are no changes in ideology. These issues also arise within a term when a newly confirmed justice joins the court mid-term and votes on a different subset of cases than her brethren. Percent conservative scores handle such instances poorly (while the measurement models discussed in the next section handle such issues well).

Using percent conservative scores as a measure of judicial ideology when trying to test theories of ideological judicial behavior strikes some scholars as circular. Segal and Cover (1989) therefore developed measures of ideology based on pre-court assessments of ideology. These scores measure the percent of paragraphs in editorials about nominations that indicate a liberal or conservative inclination in the justice. For much of the 1990s, these scores dominated the empirical literature on the Supreme Court (see, e.g., Sheehan, Mishler and Songer 1992). The scores retain the problems of percent conservative scores that arise from individuals facing dissimilar agendas, something hard to fathom given that justices are not only judged on different "votes" but that the editorial staffs (who themselves vary

ideologically) are choosing how many paragraphs to write and what to include in them. Epstein and Mershon (1996) critiques Segal and Cover scores in more detail.

For lower courts where voting data has been harder to come by, scholars have also used the party affiliation of judges or of the appointing president as a proxy for their ideological predispositions. Everyone recognizes the dangers here; such an approach for Supreme Court justices, for example, would imply that Justices Warren, Breenan, Stevens and Souter were to the right of Justice Frankfurter (and, going way back, Justice McReynolds, an arch conservative appointed by President Wilson). However, for lower courts as a whole, party explains a decent amount of variation in behavior (Pinello 1999).

2 One dimensional IRT models

Given the limitations of early approaches to measuring judicial ideology, the literature has moved decisively toward using measurement theory to estimate ideology on the court. This section covers one dimensional models, by far the most widely used models in empirical applications.

Before discussing specific approaches to measurement, it is useful to pause a moment to consider why scholars gravitate so strongly to one dimensional models. One motivation is convenience. Theoretically, equilibria in multiple dimensions are hard to come by. Statistically, identification in multiple dimensions is tricky and knowing which dimension to use and when is not easy either. A second motivation is that preferences appear, as an empirical matter, plausibly one dimensional. Grofman and Brazill (2002) find the Supreme Court is "fundamentally unidimensional" and Martin and Quinn (2002, 146) present evidence that

their one dimensional scores (which we discuss below) explain Supreme Court behavior across multiple issue areas reasonably well.

Another important motivation for one dimensional models is that sometimes the world needs to map preferences into single dimensions to make decisions (Hinich and Munger 1996). Senators voting on judicial nominees can plough through the nuances of a nominee's legal reasoning, but at the end of the day, senators may make a decisions based on whether the nominee will move the court to the left or right (Cameron and Kastellec 2015). Presidents may have similar needs when selecting nominees.

Martin and Quinn (2002) provide a prominent example of preferences based on a one dimensional item response theory (IRT) model of ideology. They update their scores annually and the scores have been used to explain opinion writing (Bonneau at al 2007), variation in ideological voting (Bartels 2009), oral arguments (Johnson, Wahlbeck and Spriggs 2006), appointments (Krehbiel 2007), congressional influence on the Court (Sala and Spriggs 2004), intercircuit conflict (Lindquist and Klein 2006) and much more.

In a classic one dimensional IRT model, the utility of actor i at time t of voting for the conservative alternative on vote v is (see, e.g., Bailey 2001)

$$u_i(\lambda_v^C) \quad = \quad -(\theta_{it} - \lambda_v^C)^2 + \eta_{iv}^C \tag{1}$$

where λ_v^C is the spatial location of the conservative alternative, θ_{it} is the ideal point of the actor at the time of proposal t and η_{iv}^C is a random shock.

The utility difference between the conservative and liberal alternatives is

$$\begin{aligned} y^*_{itv} &= -(\theta_{it} - \lambda^C_v)^2 + \eta^C_{iv} + (\theta_{it} - \lambda^L_v)^2 - \eta^L_{iv} \\ &= (\lambda^C_v - \lambda^L_v)(2\theta_{it} - (\lambda^L_v + \lambda^C_v)) + \eta^C_{iv} - \eta^L_{iv} \\ &= \alpha_v(\theta_{it} - \kappa_v) + \epsilon_{iv} \end{aligned}$$

where κ_v is the vote "cutpoint," equal to $\frac{\lambda^L_v + \lambda^C_v}{2}$ and α_v is the vote "discrimination parameter," equal to $2(\lambda^C_v - \lambda^L_v)$.[3] The error term $\epsilon_{iv} = \eta^C_{iv} - \eta^L_{iv}$ and is assumed to be a N(0, 1) random variable.

This implies that

$$\text{Prob}(y_{itv} = 1) = \Phi(\alpha_v(\theta_{it} - \kappa_v)) \tag{2}$$

The model are usually estimated with Markov Chain Monte Carlo methods.[4] MCMC methods are standard in the field and while they may sometimes appear daunting to the uninitiated, their intuition is rather simple. If we knew the vote parameters, it would be easy to use Equation 2 to estimate preferences; if we knew the preferences, it would be easy to use Equation 2 to estimate vote parameters. The magic of a MCMC approach is that under reasonable circumstances, it provides an iterative process that produces estimates from the correct distribution. Loosely speaking, if we start with some rough estimates of preferences

[3] The terminology comes from item response theory (Baker 1992). Votes for which the alternatives are relatively close (meaning $(\lambda^C_v - \lambda^L_v)$ is relatively small) will "discriminate" poorly between liberals and conservatives as the non-spatial error term will be more likely to induce actors with preferences higher than the cutpoint to vote liberally and vice versa.

[4] These models can also be estimated with EM (expectation-maximization) methods; see Bailey (2001) and Imai, Lo and Olmstead (2016).

and treat those as known, it is easy to sample from the distribution of voter parameters. If we then treat those sampled values as known, it is easy to sample from the preference distribution. If we iterate back and forth long enough, the samples we draw from both the preference and vote parameter distributions will come from the underlying distributions.

Generating ideology measures with an IRT model offers multiple benefits over percent conservative scores. First, by controlling for case characteristics (and the cutpoint in particular), IRT models deal with situations in which justices do not vote on the same cases in a given term. A justice who happened to vote on more cases with low cutpoints will have a higher percentage of conservative votes, but will not necessarily appear more conservative in IRT estimates. Second, IRT models do not weight all votes equally. Those votes that do not divide justices according to ideology get a discrimination parameter near zero which decreases their influence in the calculation of preferences. This is generally viewed as a useful way to de-emphasize "noisy" votes. In addition, MCMC IRT models produce measures of uncertainty, which are typically absent from percent conservative scores.

Many one dimensional IRT models require no coding of cases as liberal or conservative. Suppose the dependent variable is coded as 1 if a justice votes with the majority and 0 otherwise. If conservatives are in the majority on a case and they are opposed by liberals, the case discrimination parameter (α_v) will be positive, indicating conservatives are more likely to vote with the majority on that case. If liberals are in the majority on a case and they are opposed by conservatives, the case discrimination parameter (α_v) will be negative, indicating liberals are more likely to vote with the majority on that case. A discrimination parameter near zero indicates that each side had a mix of liberals and conservatives. Simplifying only

a little, automated coding says a decision is liberal if liberals voted for it and conservatives voted against and vice versa. Fischman and Law (2009) call this the agnostic approach to dealing with ideological polarity of votes.

Such automated coding of case polarity is not an unmitigated positive feature, however. Bonica and Woodruff (2014, 9) uncovered a fascinating example in which ideological polarity flummoxes unguided IRT models. They used an IRT model to scale nine state supreme courts as part of the process of validating their campaign finance based measures (which we cover below). They found that their IRT estimate of Alabama Chief Justice Roy Moore suggested he was the most liberal judge on the Alabama court. In fact, Moore is (justifiably) known as one of the most conservative judges in the United States. It appears that the IRT model in which ideological direction is not imposed interpreted votes on which Moore voted against other conservative judges as evidence that Moore was on the left, rather than the right.[5]

Many IRT models are dynamic, which, in this context, means that they allow the preferences for individual justices to evolve over time. They find considerable evidence that at least some justices change over time (Martin and Quinn 2002; Bailey 2007, 2013).[6]

Martin and Quinn generate dynamic preference estimates by including a Bayesian prior

[5] For the U.S. Supreme Court, *U.S. v. Comstock* (2009) provides another interesting example of the dangers of the agnostic approach. In *Comstock*, seven justices voted to uphold a federal law on sex offenders. Justices Scalia and Thomas voted to strike the law on the grounds that it was not effectuating an enumerated power. An agnostic approach would comfortably code this as a liberal decision as the more liberal justices lined up against the most conservative justices and while their desire to limit federal power can be interpreted as conservative, Scalia and Thomas were voting to strike a law that increased penalties on sex offenders, which is generally viewed as a liberal outcome. Using a model in which liberalism what liberals do can mask some interesting other elements of decision-making; in this case, Scalia and Thomas were willing to free individuals for whom they had little sympathy in order to promote a legal principle.

[6] Many view this as a distinctive feature of the Supreme Court relative to Congress where Poole and Rosenthal have found that most members have fixed preferences over time. However, Bailey (2007) applies the techniques discussed below to Congress and finds much more change among members of Congress. Such change often corresponds to common sense: Senator Hollings (D, SC), for example, voted against the Civil Rights Act but for the Civil Rights Act of 1991.

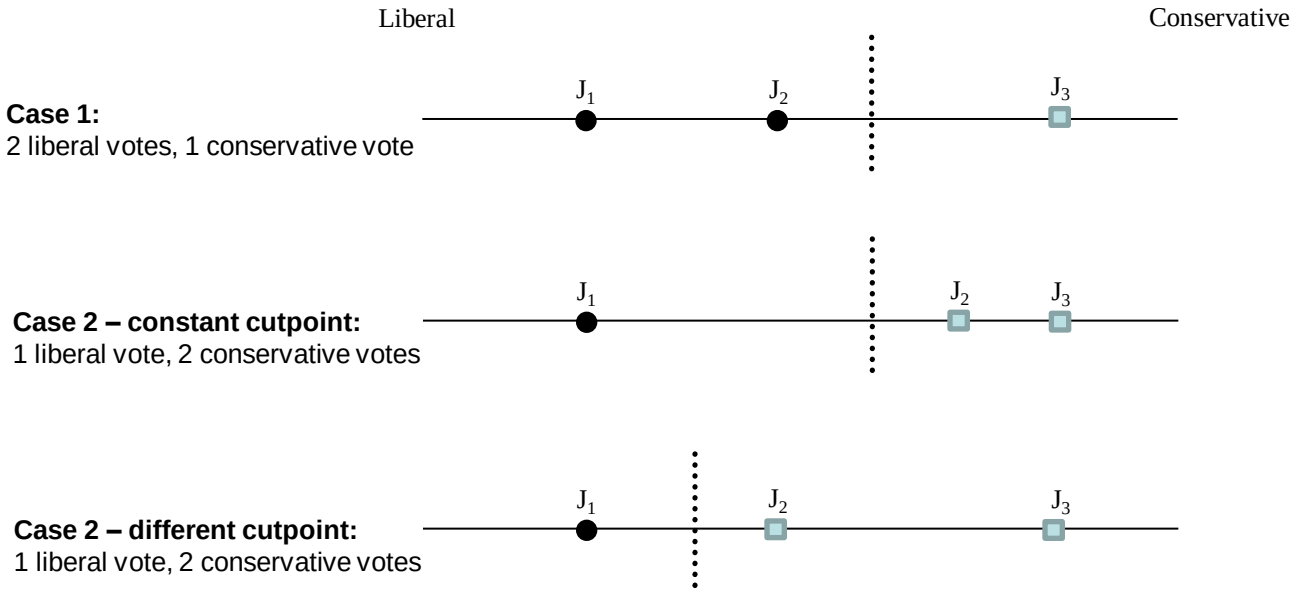

Figure 2: Difficulty in Identifying Preference Change or Cutpoint Change

that shrinks justices' ideal point estimates in a given term toward their preference estimates in previous terms. Practically, this means that if two justices vote conservatively a similar percentage of time in a given term, the one who was more conservative in the previous term will be estimated as more conservative.[7] A challenge for this approach is tuning the model by selecting an appropriate smoothing parameter. This is more of an art, than a science (Martin and Quinn 2002, 147).

One concern about Martin and Quinn scores is whether they are comparable across time.

[7] This feature allows estimated preferences to change in a more flexible way than in polynomial approaches, approaches that are frequently used in the literature (Bailey 2007; Poole and Rosenthal 1997). Bailey (2013) more or less follows Martin and Quinn in incorporating this flexible approach to preference evolution. Bailey and Maltzman (2011, chapter 6) estimate preferences separately in the period just before and just after an election in order to assess separation of powers shifts related shifts in Court behavior in order to side-step limits of preference estimates produced via polynomial function.

Many applications that use Martin and Quinn scores use them because they are "on a comparable scale over time" (Martin and Quinn 2007, 366). Such statements are true only if "the distribution of case characteristics is constant over time" (Ho and Quinn 2010, 845). Figure 2 illustrates the role this assumption plays for Martin and Quinn scores. The top row shows ideal points of three justices on "Case 1," a case on which two of the justices voted liberally and one voted conservatively. Suppose we assume that justices' ideal points can vary over time (as most work on the Court does) and consider possible ideal points on "Case 2," a case on which one of the justices voted liberally and two voted conservatively. In the first scenario, the case cutpoints on Case 2 is the same as for Case 1 and this means that justice 2 has moved to the right. However, the second scenario shows a situation in which the case cutpoint has moved left and justice 2 has not moved at all. Both scenarios for Case 2 are logically possible and, indeed, highly plausible as we could easily imagine cases that have similar ideological cutpoints as earlier cases and cases that have different cutpoints than earlier cases. For measurement, however, deciding which scenario is most accurate is crucial. Looking at the vote tally on Case 2 of two conservatives versus one liberal provides no guidance and we must rely on some external assumption or information. Martin and Quinn essentially assume that scenario 1 is true.[8]

It is possible that the Martin and Quinn approach could mistake a shift in case cutpoints for a shift in justice ideal points. Figure 3 shows Martin and Quinn estimates for the Court

[8] More precisely, their approach assumes that the distribution of votes is the same across terms. An test-grading analogy is useful here. Suppose an instructor taught the same subject two years in a row and that students scored higher in the second year. Even with an IRT model, the instructor could not be sure whether the students got better or the questions got harder. If the instructor were to use some questions in both years (and was able to prevent the test questions from leaking out) then he or she could figure out what happened. If the students in year two scored higher on the test questions asked both years, then the students got better. If the students in year two scored worse, then the other questions were easier in year two. The bridge estimates discussed below use essentially this logic.

median over time. Higher values indicate a more conservative Court median. The dashed line is the Martin and Quinn median as estimated in early 2013 (which is similar to estimates from earlier years). The solid line is from the Martin and Quinn estimates as revised during 2013 (which resemble estimates generated in 2014).

One discordant note in this series is that both Martin and Quinn estimates imply that the Court median was at one of its historically conservative peaks in 1973. This is hard to square with the fact that during this period the Court was generally considered rather liberal and produced two famously liberal (and important) decisions: in *Roe v. Wade* (1973) the Court said that there is a constitutional right to abortion and in *Furman v. Georgia* (1972) the Court imposed a nationwide moratorium on the death penalty. The Martin and Quinn scores also have a dramatic move to the left from 1973 to 1981, something inconsistent with conventional views on the Burger Court and other scaling efforts (Grofman and Brazil 2002; Bailey 2007, 2013).

Another striking aspect of Martin and Quinn scores is the drastic changes that occurred when they revised their scores in 2013. The sharp rise of the dashed line after 2008 indicates that the early 2013 Martin and Quinn scores suggested a precipitous shift to the right by the recent Roberts Court. In fact, these scores suggested that the Roberts Court was uniquely conservative in modern history. The media eagerly responded with headlines like "Supreme Court May Be Most Conservative in Modern History" (Silver 2012).

Martin and Quinn revised their scores in 2013, producing the darker line that shoots downward after 2010 in Figure 4. Whereas their earlier estimate suggested the Roberts Court is remarkably conservative, their revised estimates suggest the Roberts Court is more

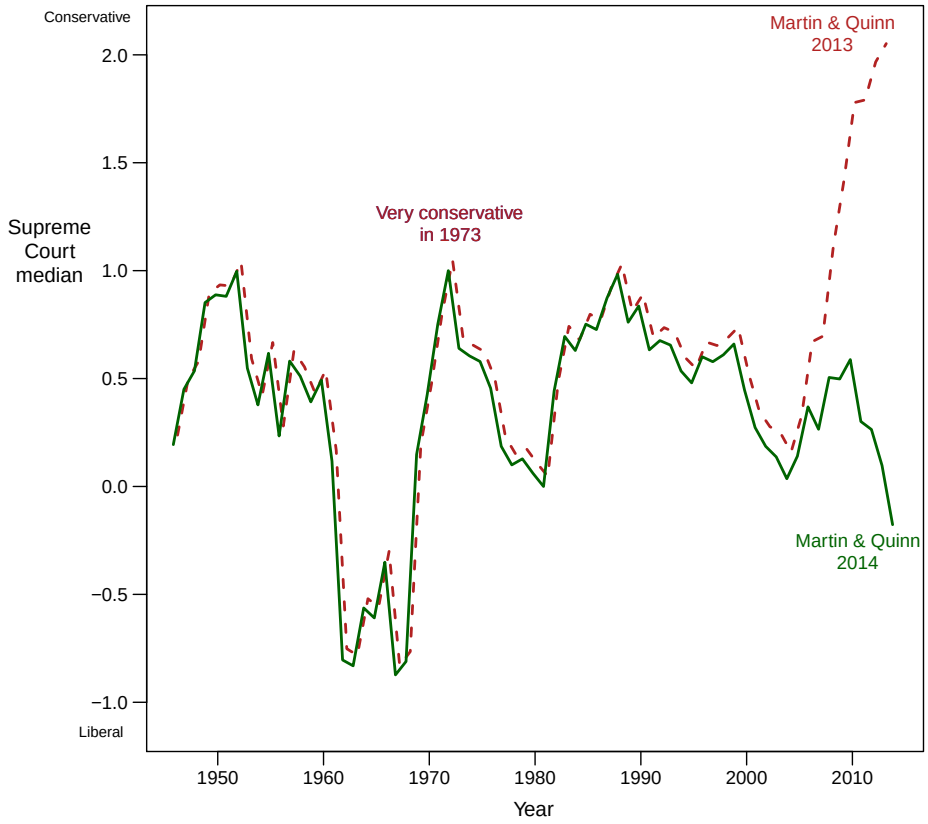

Figure 3: Martin and Quinn Estimates of Supreme Court Median over Time

liberal than any court since 1969.[9]

The key input driving the revised Martin and Quinn estimates to the left appears to be the increase in liberal decisions by the court. Is this trend, which has been onging from 2011 to 2015, strong enough to justify equating the Roberts Court to the Warren Court? Many are skeptical. Several scholars have, for reasons consistent with discussion earlier in this paper, argued that the case mix facing the court has likely shifted, which means that an increase in liberal voting should not necessarily reflect a concomitant shift to the left in the ideology of justices. Hasen (2015) points out that "liberals have wanted to stay out of the

[9] Kevin Quinn indicated in personal correspondence that the revision dealt with the fact that that from 2009 to to the first release of the 2013, their measures were based on data that included unanimous cases from 2009 onward. It is surprising that inclusion of unanimous votes had any effect, let alone such a large effect because IRT models are designed to not be moved by inclusion of unanimous votes (which contain little or no information, depending on priors imposed with regard to vote cutpoints) and given that the liberal trend observed in non-unanimous cases also existed in unanimous cases.

Court in many classes of cases (e.g., voting rights or campaign finance cases) because they know where five votes are." Fishkin and Franklin (2015) make the case with the following hypotheticals

> Imagine a tale of two Supreme Court terms. In the first, at Time One, the Court decides (a) yes, affirmative action is constitutionally required, (b) yes, the death penalty is unconstitutional, and (c) no, the government is not obligated to pay compensation to the descendants of former slaves. Two liberal decisions, one conservative decision, for a liberal rate of 66.66667%. In a term some decades later, at Time Two, the Court (having by the way overturned the first two decisions from Time One long ago) decides the following: (a) no, affirmative action is not even constitutionally permitted, but is prohibited by the Fourteenth Amendment (b) yes, the death penalty by the method of burning at the stake raises some Eighth Amendment problems, and (c) yes, despite the rulings of some lower courts, currently enslaved people can indeed bring claims under the Thirteenth Amendment in narrow circumstances. Between Time One and Time Two, the Court moved far to the right. Yet that liberal percentage remains precisely 66.66667%.... The difference is not the answers. The difference is the questions.

The inability of IRT models based on voting data alone to estimate preference change over time means that we must add additional information to the models if we wish to create credible cross-time measures. Bailey (2013) explicitly responded to Martin and Quinn's initial evidence that the Supreme Court was moving dramatically to the right by providing 'bridging estimates' that coupled IRT measurement theory with additional information that

pinned down preference changes over time (see also Bailey and Chang 2001; Bailey 2007). The analogy is to test standardization. A test score based on one set of questions is hard to compare to another score based on a different set of questions. If we gave identical tests we could compare scores, of course, but that is infeasible for standardized tests that cannot simply give identical tests year in and year out. Some questions can overlap, though, and this is enough to identify relative test performance as the overlapping questions provide information about relative ability across tests. For the Court case, we will identify certain "questions" (cases) that have been "asked" (voted on) over time. We will also be able to identify cases that are to the left or right of previous cases, information that also helps pin down agenda and preference change over time.

These bridging models used two sources of bridging information. One is the positions taken by justices on cases decided by earlier courts. It is relatively common for justices to state a clear position about an earlier decision. For example, in *FEC v. Wisconsin Right to Life* (2007) Scalia stated that *Austin v. Michigan Chamber of Commerce* (1990) was "wrongly decided"; in *Allegheny v. ACLU* (1989) Justice Kennedy wrote "I accept and indeed approve both the holding and the reasoning of Chief Justice Burger's opinion in *Lynch* [*Lynch v. Donnelly*] (1984)]." The second source of bridging information is about the relative position of cases over time. It is not unusual for case law to evolve in an ideologically understandable way. For example, the Court held in *Miller v. Alabama* (2011) that juveniles convicted of murder cannot be subject to mandatory life sentences. The cutpoint of this case was clearly to the left of *Graham v. Florida* (2010) in which the Court ruled against mandatory life sentences for juveniles for crimes other than murder. To see this, consider

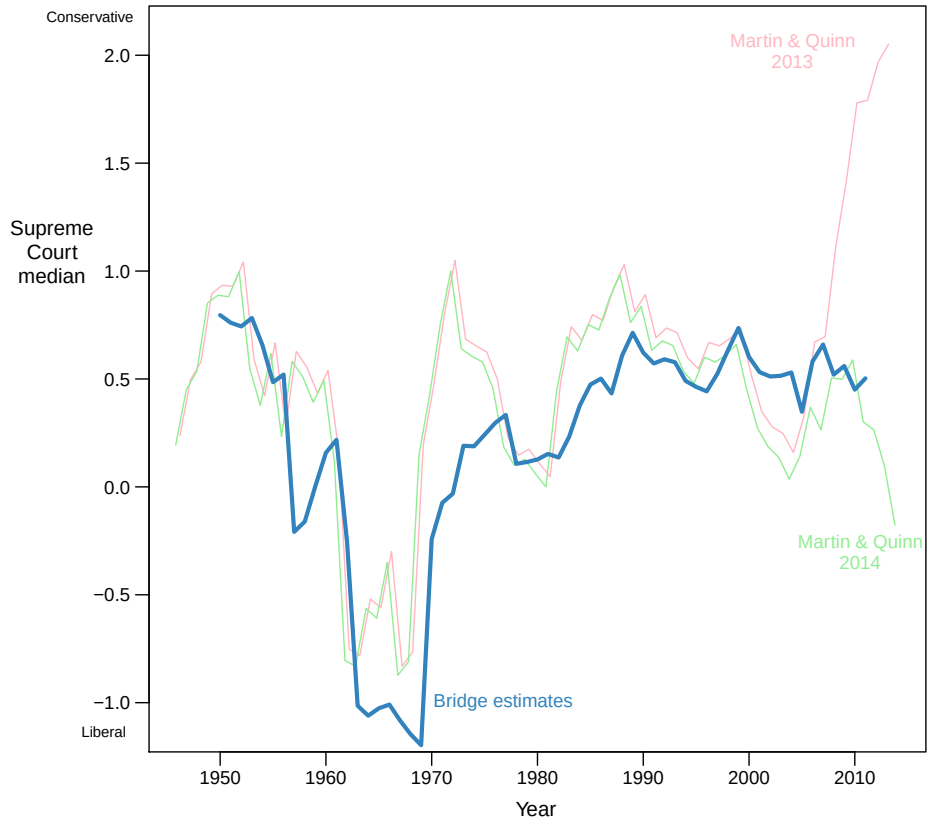

Figure 4: Bridging Estimates, Compared to Martin and Quinn Scores

the vote of a liberal on *Graham*. Does that liberal vote imply support for the liberal position on *Miller*? It does, as voting against mandatory life sentences for murder implies opposition to mandatory life sentences for lesser crimes.

Figure 4 adds the Bailey (2013) bridging estimates of the Court median over time to a figure containing the Martin and Quinn scores. As with the Martin and Quinn scores, the median nose dives under the Warren Court and bounces aggressively back in the early Nixon years. The recent court has been bouncing around a moderate conservative norm since the mid-1980s.

The bridging scores differ markedly in several respects with the Martin and Quinn scores. First, while the bridging scores do indicate that the Burger Court moved right, they do not

imply the court was at a conservative peak in 1973, which is what the Martin and Quinn scores suggest. Second, the bridging scores show no sign of the dramatic shift to the left for the Burger Court after 1974, again in contrast to the Martin and Quinn scores.

Finally, the bridging scores indicate that the Roberts Court has been conservative, in contrast to the pre-revision Martin and Quinn scores that suggested the court was extremely conservative and in contrast to the post-revision Martin and Quinn scores that suggest the court is as liberal as it was at the end of the Warren Court. Since the bridge scores from Bailey (2013) go until 2011 only, we do not know if the liberalism Martin and Quinn identify in recent years will show up in a model designed to measure change over time. Updated bridge scores similar to Bailey (2013) are expected in 2016.

The reliance of the bridging approach on external data is both a strength and a weakness. It is a strength in the sense that if anyone has specific reasons to justify thinking that the Court has moved to the right or left, this can be incorporated into the estimation. That is, if someone believes the Court has become more anti-accused and can identify either a case that is clearly to the right of a previous case (based on the substance of a ruling) or finds instances of justices critiquing earlier liberal opinions, then this information can be incorporated and used to help pin down relative movement over time. The use of this data also creates challenges, including not only the effort of identifying such external information, but also the possibility for subjectivity to enter in the collection or coding of data.

3 Inter-institutional Measures of Ideology

A related, but more challenging issue is creating ideology measures that are comparable across institutions. This is important for many research agendas, including those dealing with separation of powers. We want to know if, when and how Congress and the President affect the courts and to assess such questions systematically we need some way to characterize what actors in these various institutions want in a comparable way.

Early efforts essentially equated percent liberal scores across institutions. Segal (1997) assumed that preference measures for justices (based on Segal-Cover scores) were directly comparable to ADA scores. Moraski and Shipan (1999) assumed that percent liberal scores for justices were directly comparable to adjusted ADA scores from Groseclose, Levitt, and Snyder (1999). These assumptions are, tacitly, assumptions that the distribution of votes facing Congress and Court is identical, an unlikely possibility.

Bailey and Chang (2001) and Bailey (2007; 2013) applied bridging logic to the question, using positions taken by Presidents and members of Congress on Supreme Court cases. There are multiple sources of such observations, including amicus filings by the Solicitor General (who reports to the president) and members of Congress, public pronouncements on specific cases, and votes and cosponsorhip in Congress on bills that directly relate to cases considered by the court. By including these observations in a single IRT model, the preference estimates of actors are measured simultaneously and on the same scale. These papers also include votes in Congress on matters relating the main issues on the agenda facing the court (criminal procedure, civil rights, First Amendment, due process, privacy,

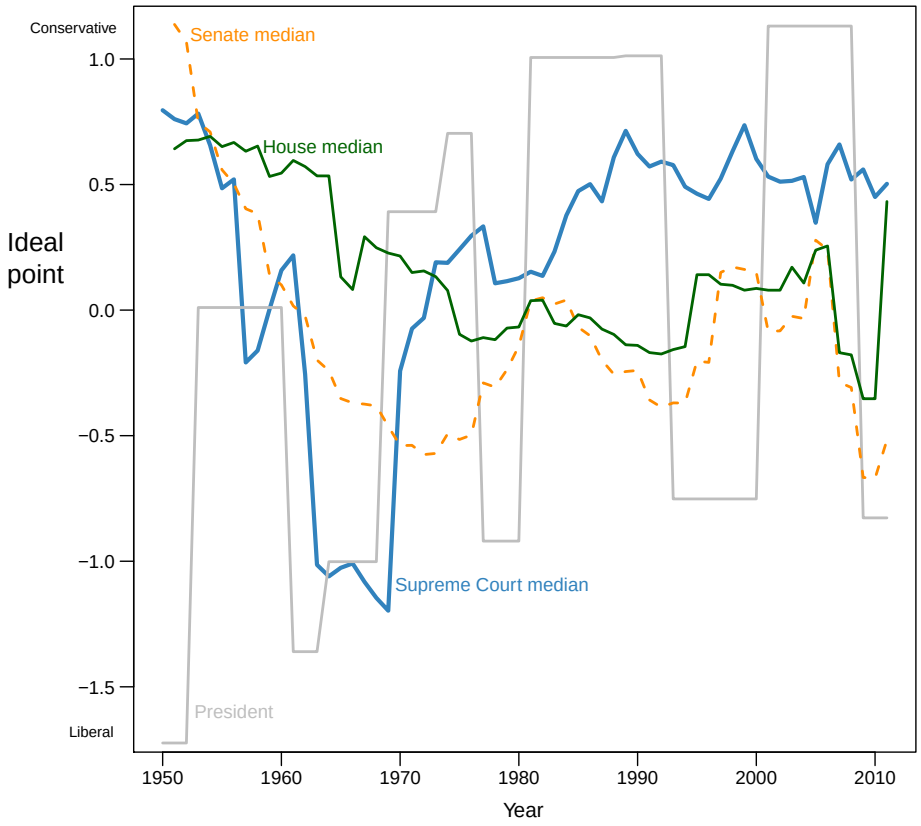

Figure 5: Inter-Institutional Bridging Estimates

and appointments to the courts).[10]

Figure 5 shows the ideology of the medians of the Supreme Court, the House, the Senate along with the ideology of the president from 1950 to 2011 as generated by Bailey (2013). The president bounces from the liberal end of the spectrum to the conservative end of the spectrum. The House median was more conservative than the court median until the mid-1970s and then was more generally more liberal. Only with the 2010 election, did the House move to the right of the court. The Senate median has generally been to the left of the House median and remained to the left of the court median even after the 2010 election. Interestingly, the court median has generally been in the "pareto set" of the elected branches

[10]This limitation is implemented in order to be as consistent with the assumption of unidimensionality. Cases related to Indian affairs, unions, economic activity, judicial power and federal taxation are excluded from the Supreme Court vote data and the congressional data includes only votes related to criminal procedure, civil rights, First Amendment, due process, privacy, and appointments to the courts.

as for most years there has been at least one elected actor to the right and one to the left. Exceptions include a period after President Nixon's election when the court was to the left of all other actors, a period under President Carter when the court was slightly to the right of Congress, the entire Clinton presidency when the Court was to the right of Congress and the president, and the first two years of the Obama presidency.

An even more challenging inter-institutional challenge is linking ideology measures of lower court judges to Congress. Data for individual judges is rather sparse (due in part to the large number of unanimous cases) and the opportunity to link such votes to positions take by members of Congress is almost nonexistent.

Two papers, taken together, use theoretical models about the appointment processes to scale preference measures across institutions into the first dimension of Poole and Rosenthal's (1997) Common Space. First, motivated by the norm of senatorial courtesy in lower court appointments, Giles, Hettinger and Peppers (2001) assume that judges appointed from a state where at least one senator is of the same party as the president will have the same Common Space score of the average of senators in the president's party from that state. For judges appointed from states with no senator from the president's party, the judge is assumed to have the Common Space score of the nominating president.

Epstein et al (2007) add Supreme Court justices to the Common Score space by mapping of Martin and Quinn scores onto the first dimension Poole and Rosenthal common space scores based on the predictions of the Moraski and Shipan (1999) model of Supreme Court nominations. By assuming that the Moraski and Shipan (1999) model characterizes Supreme Court nominations, they can translate Poole and Rosenthal Common Space scores of the

president into ideological scores for 15 justices who were nominated in so-called unconstrained regimes in which the president could nominate someone with his ideology. They then run a regression in which the Poole and Rosenthal Common Space score for these 15 justices is the dependent variable and their Martin and Quinn score in their first year of service is the independent variable.[11] Using the parameters from this regression, the Judicial Common Space score is the fitted value from this regression for all justices with Martin and Quinn measures of ideology.

The Judicial Common Space (JCS) scores provide plausible ideological measures in a very challenging, data poor context. We should not lose sight of the rather strong assumptions underlying them, assumptions that may be more problematic in some contexts more than others. First, the nomination models are open to question. Senatorial courtesy is real, but ignoring presidential preferences when there are senators from the president's party could be problematic in some cases, especially the 1960s when southern Democrats had very different preferences on civil rights than Presidents Kennedy and Johnson; assuming away presidential influence on these nominees seems fraught with potential problems. For example, a judge from Mississippi appointed in the 1960s would, by the Judicial Common Scores logic, be rated as having the ideology of the Democratic senators from the state (Eastland and Stennis). Doing so disregards the fact that the appointing president was much more liberal on civil rights.[12] In addition, the Moraski and Shipan nomination model is not sacrosanct and a number of scholars model the nomination quite differently, see, e.g., Cameron and Kastellec

[11] The actual dependent variable is $\tan \frac{\pi}{2} Common\ Space\ score$.

[12] Ironically, the weakness of using Poole and Rosenthal's first dimension actually helps in this particular instance as the Democratic senators are (wrongly) rated as moderates and some number of the judges appointed were actually moderate, due not to the moderation of the senators, but the moderation of the appointing president.

(2015) and Bailey and Spitzer (2015).[13]

More importantly, the first dimension of Poole and Rosenthal's Common Space scores is a very poor standard to use to assess congressional preferences on court-related matters for 1950s and 1960s. Recall that Poole and Rosenthal find a strong second dimension for Congress in this period, a dimension they label as a racial dimension. The first dimension becomes everything but race. The courts of the time were dealing very much with race and other matters that had strong political connections to race. Hence it is simply not credible to discuss court-related preferences in this time period using the first dimension of Poole Rosenthal; among (many) other examples, doing so would requires us to accept measures in which segregationist (and overtly racist) southern senators such as James Eastland (D, MS) were moderates to the left (!) of senators such as Edward Brooke (R, MA), the first popularly elected African American senator. After 1980 the assumptions of JCS scores are quite plausible, and represent a genuine contribution. However, for the 1950s and 1960s, there is reason to be cautious. Cameron, Kastellec, and Park (2013) provide evidence that Poole and Rosenthal's first dimension is particularly problematic between the *Brown* decision and 1970.

Bonica and Woodruff (2014) present a very different approach to generating comparable estimates across institutions. They use campaign finance contribution data to estimate ideology of state judges. These measures are comparable across institutions because they are simultaneously using this data to estimate ideological scores for presidents, members of Congress and many other public figures. The bridging assumption is that contributors

[13] Cameron and Kastellec show that the Common Space Scores are essentially the same whether or not one relies on the limits implied by the Moraski and Shipan model.

to state campaigns have same ideology as when the contribute to federal campaigns. If a state judge ran for election (as a judge or any other office), then he or she is scaled as any candidate based on scaling of donors (based in turn, on scaling of candidates to whom they gave money). This is 31 percent of judges. Judges who did not run for office, but who made contributions are then scaled together with all donors based on the recipients of their donations; this constitutes another 40 percent of judges. Another 24 percent of judges are scaled based on the estimated score of the their appointing governor or legislative body. In this step, Bonica and Woodruff use the modeling assumptions of Giles et al (2001) and Epstein et al (2007).[14]

Windett, Harden, and Hall (2015) combine Bonica and Woodruff's scores with IRT estimates of judicial voting behavior in all state supreme courts. They first estimate a dynamic IRT model for all judges in all states using Martin and Quinn's IRT model. Reminiscent of the Judicial Common Space scores approach, they then regress the Bonica and Woodruff scores on their IRT estimates to produce conversion parameters. Their estimates on a common scale are then the fitted values from that regression for all judges for whom they have IRT-based measures. Their estimates are based on votes, rather than contributions, and are dynamic, allowing preferences for individual judges to change over time as revealed in the IRT model. Dynamic estimates from a Martin and Quinn model with no other identification are a bit of a two-edged sword, however, as preference change and agenda change are hard to disentangle, although the use of the Bonica and Woodruff scores helps ease some of these concerns.

[14] Bonica and Woodruff also summarize and critique Brace, Langer and Hall's (2000) proxy-based approach to measuring preferences across state courts.

All inter-institutional preference measures share two challenges. One is that they require an extraordinary amount of data and hard work to merge and analyze massive amounts of data. For the work discussed, this is already water under the bridge, but going forward, new inter-institutional preference measures will continue to require substantial resources. A second concern is that joint scaling efforts depend deeply on the assumption that the two institutions are on the same scale. Lewis and Tausanovitch (2015, 17) argue that if that assumption is incorrect, the distribution of information can matter as "increasing the amount of legislator data will pull the estimates toward the legislator parameter values" and increasing the amount of court data will pull the estimates towards the court parameter values.

4 Multidimensional measurement models

So far, all models we have considered, from percent conservative to IRT models based on hundreds of thousands of observations, have been one dimensional. As we said, even though there are practical and intellectual reasons for focusing on one dimensional models, these reasons do not render the assumption of unidimensionality true. In fact, Fischman and Law (2009) and White (2005) express considerable skepticism that the Supreme Court has been unidimensional over time.

There are two problems with ignoring possible multidimensionality. First, changing dimensionality could induce appearances of preference change where they may be none. Farnsworth (2007, 1896) provides an example:

> Suppose, to take a simplified example, that Justice Kennedy tends to vote for

> the government in cases involving criminal procedure, but against the government in cases involving free speech, while Justice Rehnquist-a less libertarian sort of conservative- tends to vote for the government in both situations. (Both assumptions happen to be accurate.) Imagine that in term T, there are many criminal procedure cases (where the two Justices vote the same way) and few speech cases (where they don't). Then in term T + 1, there are lots of free speech cases. Kennedy's preferences may appear to drift to the left relative to Rehnquist's when they haven't really changed at all.

Second, the multidimensionality may in and of itself be interesting. We are quite confident that voting in Congress is largely one dimensional, but perhaps voting in Supreme Court and other courts is different, something that may shed light on important differences across very different institutions.

The literature includes three very different approaches to assessing multidimensionality. The first is based on classic multidimensional scaling. Schubert (1965; 1974) uses principle component analysis and related tools and finds two dimensions characterize judicial voting from the 1940s to the 1960s, dimensions he calls political and economic liberalism. Multi-dimensional scaling fell out of favor, in part because the technique led some back to unidimensional models (e.g. Grofman and Brazill 2002) and in part because the approach produces scales that require analysts to speculate about what the substantive underlying the scales actually refers to.

Recently, scholars have revived interest in such techniques and found some interesting patterns. Fischman and Jacobi (2015) applied multi-dimensional scaling techniques to the

Roberts Court and found that justices not only split along a standard left-right continuum (with Ginsburg, Kagan, Sotomayor, and Breyer on the left and Roberts, Alito, Scalia, and Thomas on the right), but also divided along a legalism-pragmatism dimension (with Breyer and Alito on the pragmatic side and Ginsburg, Scalia and Thomas on the legalistic side). This second dimension explains less variation, but can be important, especially on cases involving the Confrontation Clause, the right to a jury trial, and the Fourth Amendment. Edelman, Klein and Lindquist (2012) used multi-dimensional scaling techniques to help uncover when the court acts consensually, finding that ideology is not a sufficient explanation for consensual behavior on the court.

Another way to explore multidimensionality is to model the possibility that justices could value legal principles and the relevance of these principles could ebb and flow across cases. For example, if it is indeed the case that Kennedy is distinctively protective of free speech claims, this could not only have the measurement implications Farnsworth highlights, but it could also have implications for how ideology and the law matter on the Court. Bailey and Maltzman (2008, 2011) therefore add legal issues to the standard IRT model, producing a probability of voting conservatively of

$$Pr(y_{itv} = 1) \quad = \quad \Phi(\alpha_v(\theta_{it} - \kappa_v) + \pi_i Precedent_v + \delta_i DefCongress_v + \sigma_i Speech_v) \quad (3)$$

where π_i, δ_i, and σ_i are the weights justice i places on precedent, deference to Congress, and free speech. $Precedent_v$, $DefCongress_v$, and $Speech_v$ are precedent, deference to Congress, and speech variables. These variables are coded as 1 if the facts of the case are such that a conservative vote is consistent with the principle indicated. Bailey (2013) expanded non-

ideological variables to include deference to the executive branch and concern for Sixth Amendment rights. These variables do not constitute a comprehensive list of possible motivations for deviating from standard one-dimensional preferences, but they do cover some important sub-themes in constitutional law that some believe influence justices over and above policy motivations emphasized by Segal and Spaeth (1994; 1996; 2001; 2002). Law in this context is not quite some legal truth inherent in the law (which may only exist as an ideal type), but rather some motivation, rooted in legal theory or practice that induces justices to vote in a manner inconsistent with their policy preferences.

The key to identifying these models is that Bailey and Maltzman include congressional actors in the full model. The legal variables are constrained to be zero for these actors; were they not, the main effects of the legal variables would simply be absorbed by the case cutpoint variable (note that κ and the legal variables all have the same subscript). Suppose, for simplicity, that $\alpha_v = 1$ and $\pi_i = 1$ for all justices (meaning all justices care about precedent when voting). If we estimate Equation 3 using only Supreme Court justices, we could generate the same probability of conservative votes by setting $\pi_i = 0$ for all justices and $\kappa'_v = \kappa_v + Precedent_v$. If we include members of Congress, however, such an identification problem goes away as we cannot roll the effect of precedent into the case cutpoint as the case cutpoint for members of Congress will depend only on ideology.

The results suggest extensive deviations from one-dimensional policy preferences. Every justice has a significant coefficient on at least one, and usually more, legal variable. And every legal variable has multiple justices who appear influenced by the legal variable. Of the 28 justices analyzed in Bailey and Maltzman (2011), 16 were statistically significantly

influenced by First Amendment, 14 were statistically significantly influenced by precedent, and 12 were statistically significantly influenced by deference to Congress. Bailey (2013) found additional influence of the Sixth Amendment, consistent with Fischman and Jacobi's (2015) finding that a second dimension mattered for cases involving Confrontation Clause and the right to jury trials. There is strong variation over time as to what non-ideological values matter: in the fifties and sixties, justices were more inclined to defer to Congress than today. In the contemporary court, justices are more likely to be influenced by precedent than earlier justices.

There is also huge variation across justices and concepts, however, suggesting that the additional dimensions are highly idiosyncratic. This implies that while the court appointment process may be relatively predictable with regard to underlying policy preferences of justices, it appears to be more variable with regard to the additional legal influences on justices. Given the very small size of the court, this means that the court may indeed act as a wild card in the operation of the U.S. shared powers system.

Lauderdale and Clark (2012) carry this logic even further and estimate separate preferences for justices on every single case. This works because they link cases via their citation patterns, allowing the weight of preferences on one case on another case to depend on the depth of ties between the two cases. They find striking evidence of deviations from unidimensionality as the median of the court varies substantially across cases. For example, even though Justice Scalia is viewed, naturally enough, as being very conservative, there are some subsets of cases where he is the decisive vote given his particular preferences as expressed on the relevant case history. For example, Scalia may well have been the pivotal voter on

the landmark *Texas v. Johnson* case that overturned a ban on flag-burning. In a case like this, Scalia's conservativism combined with his views on the First Amendment to make him a centrist justice, as least on this case.

5 Just how conservative is the Roberts Court?

While measuring ideology may seem somewhat technical and sterile, the task is actually highly politically relevant and quite unsettled.

The political relevance is that people on both sides of the political spectrum cannot seem to agree on just what the ideology of the Roberts Court is. Many liberals and conservatives agree the court dangerously radical; they only disagree about which cliff the court is steering us off. This matters. Bartels and Johnston (2013) show that subjective ideological disagreement with the court is strongly associated with lower views of the court's legitimacy.[15]

The result is that the public can be a bit discombobulated. Hetherington and Smith (2007) found that conservatives in the early 2000s were moving against the court, despite its seeming move to the right over time. Gallup (2015) reports that the 37 percent of the public say the court is too liberal, 40 percent say it is "just about right" and 20 percent say it is too conservative. Gibson (2012, 9) found that Americans who were dissatisfied with the Court in 2011 were equally likely to say the Court was too liberal or too conservative, leading him to conclude "there is no consensus in American politics today about the ideological location of the current Supreme Court."

[15] This state of affairs has existed for some time now. Friedman (2009, 323) notes that in the early 2000s, liberal and conservative critics of the court published anti-court polemics with identical titles, *Courting Disaster* (Garbus 2002; Robertson 2004).

The liberal case that the court is conservative is based on multiple pieces of evidence. Justices appointed by Presidents Reagan, Bush I and Bush II constitute a majority on the court and have produced many landmark conservative decisions, including *Citizens United* (2010), *Hobby Lobby* (2014), *Shelby County* (2013) and *McDonald v. Chicago* (2010). And the sympathy of the Roberts Court with business interests seems undeniable, as attested by high levels of support for the Chamber of Commerce from the current justices (Epstein, Landes and Posner 2013).

Conservatives counter with signs the Roberts Court is actually liberal. Two cases are at the center of their case: *National Federation of Independent Business v. Sebelius* which largely upheld the Affordable Care Act and *Obgerfell* (2015) which held that same-sex marriage is a right guaranteed by the Constitution. Fallout from about these cases is evident in public opinion polling as Republican approval of the Court fell to 18 percent in the summer after the same-sex marriage decision (even as Democratic approval of the Court rose to 76 percent) (McCarthy 2015). The court's liberalism may be broader these two high profile decisions, however, as the percent of cases decided liberally was the highest in 2015 since 1969 (Parlapiana, Liptak and Bowers 2015).

In addition, the Court is probably not too far out of step with public opinion. Malhotra and Jessee (2013) polled Americans on specific Supreme Court cases, presenting them summaries of each position and asking which way the Court should have ruled. Of the nine decisions they polled with clear ideological valences (*Comstock*, discussed earlier, was the tenth case they polled), eight were conservative decisions (*Citizens United, Heller, Salazar, Ricci, Crawford, Baze, Parents Involved, Gonzales v. Carhart*) and respondents agreed with

the actual decision 71.6 percent of the time. It was the one liberal decision in the survey, *Hamdan*, that had the lowest level of popular support, at only 29.9 percent agreement. Given that the public had not shifted dramatically to the right (Stimson 1999, 2012) then these results were in tension with Martin and Quinn's claim as of 2013 that the current Court was the most conservative Court in the postwar era.

Given the contested interpretations of the court in the public sphere, it is natural to turn to the measurement models of the sort discussed so far in this paper. At this point, we do not have bridge estimates for the most recent court, making it difficult to come to a final view on just how liberal the court has become. Although Martin and Quinn's revised estimates suggest the court has moved firmly leftward, Jessee and Tahk's (2011, 2015) IRT model does not seem to show a shift in Kennedy's ideal point to the left, a necessary condition for the court median to shift to the left.

In the meantime, keeping in mind that the Court is not simply a little Congress helps reconcile the conflicting ideological signals emanating from the court. In Congress, if members are very conservative on voting rights, they tend to be very conservative on abortion, on foreign policy, and on welfare spending. The Court, however, has several institutional features that complicate the relatively simple ideological story that works when describing Congress. The first is that life tenure gives more justices freedom to deviate from ideological norms. Members of Congress certainly can deviate from their expected ideological position at times, but they risk defeat in primaries and marginalization within their parties. In addition, the ideological environment justices operate in is different. Yes, the role of government is an absolutely central question, but, unlike in Congress, other values are taken seriously as

well, including the role of precedent, the legitimacy of the court to overrule elected officials and more (Bailey and Maltzman 2011). In addition, agenda control in Congress is different as party leaders can chose (and possibly combine) the policy options they vote on. Certainly Supreme Court justices control their agenda too, but lower court splits may force their hand and they generally do not have the latitude congressional leaders have to combine issues in a way that maintains coalitional cohesion.

And, perhaps above all, the small size of the court matters. While members of Congress do not lack in idiosyncracies, their electoral environment mutes some temptations to deviate from ideological expectations. And, many other surprising ideological choices simply wash out in the large chambers, as it is seldom the case that a single individual member of Congress exerts virtual control over the legislative outcome. On the Supreme Court, on the other hand, there are fewer institutional mechanisms to constrain justices and fewer decision-makers to wash out the idiosyncracies that persist. This means that the court is more prone to be driven by idiosyncratic positions that don't necessarily map easily onto the main political ideological space. The result is that the Roberts Court was (until the death of Justice Scalia) dominated by Justice Kennedy. Kennedy is a moderate conservative who is liberal on some issues (such as same-sex marriage) and conservative on other issues (such as campaign finance). The result is that even though the court is generally moderately conservative, it can produce a conservative blockbuster like *Shelby County* (2013) on one day and liberal blockbusters like *Windsor* (2013) and *Hollingsworth* (2013) on (literally) the next day.[16]

[16] The Rehnquist Court also exhibited some ideological incongruities. Bryden (1992, 74) noted that even as the Rehnquist Court was to the right of the Warren and Burger Courts, it was "rarely or never more conservative than John Q. Public's opinion."

6 Conclusion

Measuring the ideology of courts is inevitable. Research demands these measures for broad array of important agendas. And, real-world political actors often need to summarize court behavior as they consider nominees or consider possible institutional changes that constrain or empower courts.

The last two decades have witnessed remarkable progress in measuring ideology on the courts. At one point, scholars were forced to use measures they fully recognized were deeply flawed. Everybody recognized that even though percent conservative scores provide a rough and ready measure of preferences, they do little to control for agenda and are poorly suited to making cross-time or inter-institutional comparisons. Since then, the literature has embraced model-based measurement models and has produced measures that summarize ideology not only on the court in isolation, but for the court compared to previous courts, Congress, the president, lower courts and even state political actors.

For all our progress, however, we cannot forget that measuring ideology is tricky. Measurement models provide theoretically compelling frameworks for tackling hard problems, but they nonetheless require assumptions, assumptions that might be more problematic for some research agendas than others. The many challenges include the following:

- Cross-time comparisons require information pinning down the movement of agenda over time. Incorporating such data is time intensive and introduces the possibility of subjective coding.

- Cross institutional comparisons require information pining down the preferences of

actors across institutions. Any measure that uses Poole and Rosenthal first dimensional preferences from before 1970 implicitly assumes that southern segregationists were "moderate," which is wildly unrealistic with regard to their position on civil rights and other issues that dominated the court's docket in that era.

- Models with multiple dimensions typically require additional time-consuming data collect and are statistically complicated.

However much progress we make measuring ideology on the court, the reality is that ideology on the court will always differ, at least somewhat, from ideology among elected officials. Justices have more latitude to define ideology in highly personal ways and do their work where factors such as *stare decisis* can inject perturbations in the ideological space seldom seen in the ideology of elected officials. Given the small size of the court, small idiosyncratic deviations from ideological expectations can result in major surprises even when we have a good sense of the general ideological orientation of the justices.

Going forward, we will have more data and bigger computers, but we will also need more models and better theory to truly understand court behavior. The challenges we see in estimating ideology for the Roberts Court will persist, which should not discourage us, but rather give us confidence that courts will continue to be fascinating and unique players in the U.S. political system.

Bibliography

Bailey, Michael A. 2013. Is Today's Court the Most Conservative in Sixty Years? Challenges and Opportunities in Measuring Judicial Preferences. *Journal of Politics* 75, 3: 821-834.

Bailey, Michael A. 2007. Comparable Preference Estimates Across Time and Institutions for the Court, Congress and Presidency. *American Journal of Political Science* 51, 3: 433-448.

Bailey, Michael A. 2001. Ideal Point Estimation with a Small Number of Votes: A Random Effects Approach. *Political Analysis* 9, 3: 192-210.

Bailey, Michael A. and Matthew L. Spitzer. 2015. Appointing Extremists. Research Paper No. 15-04; Northwestern Public Law Research Paper No. 15-07. Available at SSRN: http://ssrn.com/abstract=2568129.

Bailey, Michael A. and Forrest Maltzman. 2008. Does Legal Doctrine Matter Unpacking Law and Policy Preferences on the U.S. Supreme Court. *American Political Science Review* 102, 3: 369- 384.

Bailey, Michael A. and Forrest Maltzman. 2011. *The Constrained Court: Law, Politics and the Decisions Justices Make*. Princeton University Press.

Bailey, Michael and Kelly Chang. 2001. Comparing Presidents, Senators, and Justices: Inter-institutional Preference Estimation. *Journal of Law, Economics and Organization* 17: 477-506.

Bartels, Brandon L. 2009. The Constraining Capacity of Legal Doctrine on the U.S. Supreme Court. *American Political Science Review* 103, 3: 474-95.

Bartels, Brandon L. and Christopher D. Johnston. 2013. On the Ideological Foundations of Supreme Court Legitimacy in the American Public *American Journal of Political Science* 57, 1: 184-99.

Baum, Lawrence. 1988. Measuring Policy Change in the U.S. Supreme Court. *American Political Science Review* 82, 3: 905-912.

Bhatia, Kedar. 2014. Final Stat Pack for October Term 2014. *SCOTUSblog* `http://www.scotusblog.com/2015/06/final-stat-pack-for-october-term-2014/`.

Bonica, Adam and Michael J. Woodruff. 2014. A Common-Space Measure of State Supreme Court Ideology. *The Journal of Law, Economics, and Organization*.

Bonica, Adam. 2014. Mapping the Ideological Marketplace. *American Journal of Political Science* 58, 2: 367 - 386

Bonneau, Chris W., Thomas H. Hammond, Forrest Maltzman and Paul J. Wahlbeck. 2007. Agenda Control, the Median Justice, and the Majority Opinion on the U.S. Supreme Court. *American Journal of Political Science* 51, 4: 890-905.

Brace, Paul, Laura Langer, and Melinda Gann Hall. 2000. Measuring Preferences of State Supreme Court Judges. *Journal of Politics* 62, 2: 387 - 413.

Bryden, David P. 1992. Is the Rehnquist Court Conservative? *Public Interest* 109(Fall): 73-88.

Cameron, Charles M. and Kastellec, Jonathan P. 2015. Are Supreme Court Nominations a Move-the-Median Game? Available at SSRN: http://ssrn.com/abstract=2594947.

Cameron, Charles M., Kastellec, Jonathan P. and Jee-Kwang Park. 2013. Voting for Justices: Change and Continuity in Confirmation Voting 1937–2010. *The Journal of Politics*

75, 2: 283299.
Chiou, Fang-Yi and Kosuke Imai. 2008. Nonparametric and Semiparametric Identification Problems in the Statistical Analysis of Roll Call Data, Princeton Working Paper.
Edelman, Paul H., David E. Klein, and Stefanie A. Lindquist. 2012. Consensus, Disorder, and Ideology on the Supreme Court. *Journal of Empirical Legal Studies* 9, 1: 129148.
Epstein, Lee, William M. Landes, and Richard A. Posner. 2013. How Business Fares in the Supreme Court. *Minnesota Law Review* 97: 1431 - 1472.
Epstein, Lee and Jeffrey A. Segal. 2006. Trumping the First Amendment. *Journal of Law & Policy* 21: 81-121.
Epstein, Lee and Carol Mershon. 1996. Measuring Political Preferences. *American Journal of Political Science* 40, 1: 261-294.
Epstein, Lee, Andrew Martin, Jeffrey A. Segal and Chad Westerland. 2007. The Judicial Common Space. *Journal of Law, Economics and Organization* 23, 2: 303-325.
Farnsworth, Ward. 2007. The Use and Limits of Martin-Quinn Scores to Assess Supreme Court Justices, with Special Attention to the Problem of Ideological Drift. *Northwestern University Law Review* 101, 4: 1891-1904.
Fishkin, Joey and Cary Franklin. 2015. What was the question? *Balkinization* (June 23) `http://balkin.blogspot.com/2015/06/what-was-the-question.html`
Fischman, Joshua B. and Tonja Jacobi. 2015. The Second Dimension of the Supreme Court. *William & Mary Law Review*. Forthcoming.
Fischman, Joshua B. and David S. Law. 2009. What Is Judicial Ideology, and How Should We Measure It? *Washington University Journal of Law and Policy* 29: 133- 213.
Friedman, Barry. 2009. *The Will of the People: How Public Opinion Has Influenced the Supreme Court and Shaped the Meaning of the Constitution.* New York: Farrar, Straus and Giroux.
Gallup. 2015. Supreme Court. In Depth: Topics A to Z. `http://www.gallup.com/poll/4732/Supreme-Court.aspx`.
Garbus, Martin. 2002. *Courting Disaster: The Supreme Court and the Unmaking of American Law.* New York: Times Books.
Gibson, James. 2012. Public Reverence for the United States Supreme Court: Is the Court Invincible? Manuscript, Washington University in St. Louis.
Giles, Micheal W., Virgina A. Hettinger, and Todd Peppers. 2001. Picking Federal Judges: A Note on Policy and Partisan Selection Agendas. *Political Research Quarterly* 54: 62341.
Grofman, Bernard and Timothy J. Brazill. 2002. Identifying the Median Justice on the Supreme Court Through Multidimensional Scaling: Analysis of the Natural Courts 1953-1991. *Public Choice* 112: 55-79.
Groseclose, Timothy, Steven Levitt and James Snyder. 1999. Comparing Interest Group Scores across Time and Chambers: Adjusted ADA Scores for the U.S. Congress. *American Political Science Review* 93: 33-50.
Hansford, Thomas G. 2012. Using the Amici Network to Measure the Ex Ante Ideological Loading of Supreme Court Cases. Presented at Midwest Political Science Association Annual Meeting.
Harvey, Anna and Michael Woodruff. 2011. Confirmation Bias in the United States Supreme Court Judicial Database. *Journal of Law, Economics and Organization.*

Hasen, Rick. 2015. Whats the Significance of the Roberts Courts Relatively Liberal Term? *Election Law Blog* (June 23) http://electionlawblog.org/?p=73696

Hetherington, Marc J., and Joseph L. Smith. 2007. Issue Preferences and Evaluations of the U.S. Supreme Court. *Public Opinion Quarterly* 71, 1: 40-66.

Hinich, Melvin and Michael C. Munger. 1996. *Ideology and the Theory of Political Choice*. Ann Arbor: University of Michigan Press.

Ho, Daniel E. and Kevin M. Quinn. 2010. How Not to Lie with Judicial Votes. *California Law Review* 98: 813- 876.

Imai, Kosuke, James Lo and Jonathan Olmsted. 2016. Fast Estimation of Ideal Points with Massive Data. Manuscript, Princeton University.

Jessee, Stephen and Alexander Tahk. 2011. What Can We Learn About the Ideology of New Supreme Court Justices? *PS: Political Science and Politics* 44, 3: 524-529.

Jessee, Stephen and Alexander Tahk. 2015. Supreme Court Ideology Project. http://sct.tahk.us/.

Johnson, Timothy R, Paul J. Wahlbeck and James F. Spriggs II. 2006. The Influence of Oral Arguments on the U.S. Supreme Court. *American Political Science Review* 100, 1: 99-113.

Krehbiel, Keith. 2007. Supreme Court Appointments as a Move-the-Median Game. *American Journal of Political Science* 51, 2: 231-240.

Lauderdale, Benjamin E. and Tom S. Clark. 2012. The Supreme Courts Many Median Justices. *American Political Science Review* 106, 4: 847-866.

Lauderdale, Benjamin E. 2010. Unpredictable Voters in Ideal Point Estimation. *Political Analysis* 18, 2: 151-171.

Lewis, Jeffrey B. and Chris Tausanovitch. 2015. When Does Joint Scaling Allow For Direct Comparisons of Preferences? Manuscript, UCLA, Prepared for the 2015 Conference on Ideal Point Models at MIT.

Lindquist, Stefanie A. and David E. Klein. 2006. The Influence of Jurisprudential Considerations on Supreme Court Decisionmaking: A Study of Conflict Cases. *Law & Society Review* 40, 1: 135-162.

Malhotra, Neil and Stephen Jessee. 2013. Public (Mis)Perceptions of Supreme Court Ideology: A Method for Directly Comparing Citizens and Justices. *Public Opinion Quarterly* 77, 2: 619-634.

Martin, Andrew and Kevin Quinn. 2002. Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953-1999. *Political Analysis* 10, 2: 134-153.

Martin, Andrew D. and Kevin M. Quinn. 2007. Assessing Preference Change on the US Supreme Court *Journal of Law, Economics, & Organization* 23, 2: 365-385.

Martin, Andrew, Kevin Quinn and Lee Epstein. 2005. The Median Justice on the U.S. Supreme Court. *North Carolina Law Review* 83, 5: 1275-1322.

McCarthy, Justin. 2015. Disapproval of Supreme Court Edges to New High. Gallup.com at http://www.gallup.com/poll/185972/disapproval-supreme-court-edges-new-high.aspx.

Moraski, Bryon and Charles Shipan. 1999. The Politics of Supreme Court Nominations: A Theory of Institutional Constraints and Choices. *American Journal of Political Science* 43: 1069-1095.

Parlapiana, Alicia, Adam Liptak and Jeremy Bowers. 2015. The Roberts Courts Surprising Move Leftward. *New York Times - The Upshot* at `http://www.nytimes.com/interactive/2015/06/23/upshot/the-roberts-courts-surprising-move-leftward.html` (updated June 29).

Pinello, Daniel R. 1999. Linking Party to Judicial Ideology in American Courts: A Meta-analysis. *The Justice System Journal* 20, 3: 219-254.

Poole, Keith T., and Howard Rosenthal. 1997. *Congress: A Political-Economic History of Roll- Call Voting.* Oxford: Oxford University Press.

Pritchett, C. Herman. 1970. 'Tn Years of Supreme Court Voting': A Comment in Retrospect. *Social Science Quarterly* 50, 4: 983.

Pritchett, C. Herman. 1948. *The Roosevelt Court: A Study in Judicial Politics and Values, 1937-1948.* New York: Macmillan Co.

Robertson, Pat. 2004. *Courting Disaster: How the Supreme Court is Usurping the Power of Congress and the People.* Wake Forest, NC: Integrity Publishers.

Sala, Brian R. and James F. Spriggs II. 2004. Designing Tests of the Supreme Court and the Separation of Powers. *Political Research Quarterly* 57, 2: 197-208.

Schubert, Glendon A. 1965. *The Judicial Mind: the Attitudes and Ideologies of Supreme Court Justices, 1946-1963* Evanston, IL: Northwestern University Press.

Schubert, Glendon A. 1974. *The Judicial Mind Revisited* Oxford: Oxford University Press.

Segal, Jeffrey. 1997. Separation of Powers Games in the Positive Theory of Congress and Courts. *American Political Science Review* 91: 28-44.

Segal, Jeffrey A. and Albert Cover. 1989. Ideological Values and the Votes of U.S. Supreme Court Justices. *American Political Science Review* 83, 2: 557-565.

Segal, Jeffrey A. and Harold J. Spaeth. 1994. The Attitudinal Model: The Authors Respond. *Law and Courts Newsletter* 4, 1: 10.

Segal, Jeffrey A. and Harold J. Spaeth. 1996. The Influence of Stare Decisis on the Votes of the United States Supreme Court Justices. *American Journal of Political Science* 40, 4: 971-1003.

Segal, Jeffrey A. and Harold J. Spaeth. 2001. *Majority Rule or Minority Will: Adherence to Precedent on the U.S. Supreme Court.* New York: Cambridge University Press.

Segal, Jeffrey A. and Harold J. Spaeth. 2002. *The Supreme Court and the Attitudinal Model Revisited.* New York: Cambridge University Press.

Shapiro, Carolyn. 2009. Coding Complexity: Bringing Law to the Empirical Analysis of the Supreme Court. *Hasting Law Journal* 60 (3): 477 543.

Shapiro, Carolyn. 2010. The Context of Ideology: Law, Politics, and Empirical Legal Scholarship. *Missouri Law Review* 75 (1): 79 142.

Sheehan, Reginald S., William Mishler and Donald R. Songer 1992. Ideology, Status, and Differential Success of Direct Parties Before the Supreme Court. *American Political Science Review* 86, 2: 464 471.

Silver, Nate. 2012. Supreme Court May Be Most Conservative in Modern History. blog post at FiveThirtyEight [Posted March 29 to `http://fivethirtyeight.blogs.nytimes.com/2012/03/29/supreme-court-may-be-most-conservative-in-modern-history/`].

Spaeth, Harold, Lee Epstein, Andrew D. Martin, Jeffrey A. Segal, Theodore J. Ruger, and Sara C. Benesh. 2015. 2015 Supreme Court Database, Version 2015 Release 01. `http:`

//Supremecourtdatabase.org.
Solum, Lawrence B. 2014. The Positive Foundations of Formalism: False Necessity and American Legal Realism. *Harvard Law Review* 120: 2464- 2497.
Spaeth, Harold J., Lee Epstein, Andrew D. Martin, Jeffrey A. Segal, Theodore J. Ruger, and Sara C. Benesh. 2015 Supreme Court Database, Version 2015 Release 01. http://Supremecourtdatabase.org.
Stimson, James A. 1999 (2012). *Public Opinion in America: Moods, Cycles, and Swings*, 2nd Edition. Boulder, CO: Westview Press. Updated measures available at http://www.unc.edu/~cogginse/Policy_Mood.html.
White, G. Edward. 2005. Unpacking the Idea of the Judicial Center. *North Carolina Law Review* 83: 1089 - 1186.
Windett, Jason H., Jeffrey J. Harden, Matthew E.K. Hall. 2015. Estimating Dynamic Ideal Points for State Supreme Courts. *Political Analysis* 23, 3: 461-469.